Homework Assignment 1: Probability Review

Rushabh Ashok Dharia
CSCI-B555 Machine Learning

January 30, 2019

Note: Homework was discussed with Varun Miranda and Bhushan Malgaonkar

Ans 1. Container 1: 2000 Processors, 5% defective = 100 defective processors
Container 2: 500 Processors, 15% defective = 75 defective processors
Container 3: 600 Processors, 100 defective processors

a). P(C1) = P(C2) = P(C3) = 1/3
P(D/C1) = 5/100, P(D/C2) = 15/100, P(D/C3) = 1/6

$$P(\text{all 3 are defective}) = \frac{100(99)(98)}{3(2000)(1999)(1998)} + \frac{75(74)(73)}{3(500)(499)(498)} + \frac{100(99)(98)}{3(600)(599)(598)}$$
$$= 0.00263$$

□

b).

$$P(\frac{C2}{3D}) = \frac{P(C2).P(\frac{3D}{C2})}{P(3D)}$$
$$= \frac{\frac{75(74)(73)}{3(500)(499)(498)}}{0.0263}$$
$$= 0.413$$

□

Ans 2. A - P(H) = 0.75, P(T) = 0.25
B - P(H) = 0.5, P(T) = 0.5
C - P(H) = 0.25, P(T) = 0.75

a). Different Permutations

$$
\begin{aligned}
HHH &= 3*0.75*0.5*0.25\\
HHT &= 2*0.75*0.5*0.75\\
HTH &= 2*0.75*0.5*0.25\\
HTT &= 1*0.75*0.5*0.75\\
THH &= 2*0.25*0.5*0.25\\
THT &= 1*0.25*0.5*0.75\\
TTH &= 1*0.25*0.5*0.25\\
TTT &= 0*0.25*0.5*0.75
\end{aligned}
$$

$$
\begin{aligned}
E(X) &= \sum f_i x_i\\
&= 0.5*(3*0.75*0.25+2*0.75^2+2*0.75*0.25+0.75^2+2*0.25^2+0.25*0.75+0.25^2)\\
&= 0.5*(3*0.75^2+(3+2+1)*0.75*0.25+3*0.25^2\\
&= \frac{3}{2}(0.75+0.25)^2\\
&= 1.5
\end{aligned}
$$

$\square$

b).

$$
\begin{aligned}
P(C/3H) &= \frac{P(C)*P(\frac{3H}{C})}{P(3H)}\\
&= \frac{\frac{1}{3}*0.25^3*0.75^2}{\frac{1}{3}*0.25^3*0.75^2+\frac{1}{3}*0.75^3*0.25^2+\frac{1}{3}*0.5^5}\\
&= 0.1323
\end{aligned}
$$

$\square$

Ans 3. *a).* Verification

$$0 \leq \frac{\lambda^x e^{-\lambda}}{x!} \leq 1$$

$$\text{And, } \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \quad [TaylorSeries = 1 + \lambda + \frac{\lambda^2}{2!} + ... + \frac{\lambda^3}{3!} \quad = e^{\lambda}]$$

$$= e^{-\lambda} * e^{\lambda}$$

$$= 1$$

Therefore PMF is valid

□

b).

$$E(\frac{1}{1+X}) = \sum_{x=0}^{\infty} \frac{1}{1+x} P(X = x)$$

$$= \sum_{x=0}^{\infty} \frac{1}{1+x} \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{1}{1+x} \frac{\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{(x+1)!}$$

$$= \frac{e^{-\lambda}}{\lambda} \sum_{x=0}^{\infty} \frac{\lambda^{x+1}}{(x+1)!} [\text{put } y = x + 1.\text{Therefore when } x = 0, y = 1 \text{ and when } x = \infty y = \infty]$$

$$= \frac{e^{-\lambda}}{\lambda} [\sum_{y=1}^{\infty} \frac{\lambda^y}{y!}]$$

For y=0 $\dfrac{\lambda^0}{0!} = 1$

Therefore, adding + 1 and -1 to the above equation, we get

$$= \frac{e^{-\lambda}}{\lambda} [\sum_{y=0}^{\infty} \frac{\lambda^y}{y!} - 1]$$

$$= \frac{e^{-\lambda}}{\lambda} (e^{\lambda} - 1) \text{ [Using Taylor series]}$$

$$= \frac{1 - e^{-\lambda}}{\lambda}$$

□

Ans 4. *a).*

$$\begin{aligned}\text{Arc Length } S = r\theta &= X\\ \text{Area} = \frac{r^2\theta}{2} &= Y\\ 2\pi r &= 1\\ r = \frac{1}{2\pi}\,\frac{X}{Y} &= \frac{r\theta}{\frac{r^2\theta}{2}}\\ \frac{X}{Y} = \frac{2}{r} &= 2 * 2\pi\\ Y &= \frac{X}{4\pi}\end{aligned}$$

$\square$

b).

$$\begin{aligned}F_Y(y) &= P(Y \leq y)\\ &= P(\frac{X}{4\pi})\\ &= P(X \leq 4\pi y)\end{aligned}$$

Therefore,

$$F_Y(y) = \begin{cases} 0 & y \leq 0 \\ 4\pi y & 0 \leq y \leq \frac{1}{4\pi} \end{cases}$$

$\square$

c).

$$f_Y(y) = \frac{d}{dx}F_Y(y)$$

$$f_Y(y) = \begin{cases} 4\pi & 0 \leq y \leq \frac{1}{4\pi} \\ 0 & \text{Otherwise} \end{cases}$$

$\square$

d).

$$\begin{aligned}E[Y] &= \int_{-\infty}^{\infty} g(y).f(y)dy\\ &= \int_0^{\frac{1}{4\pi}} y.4\pi dy\\ &= 4\pi \int_0^{\frac{1}{4\pi}} y dy\\ &= \frac{1}{8\pi}\end{aligned}$$

$\square$

Ans 5. *a).*

$$
\begin{aligned}
E[N] &= \sum_{n \in N} n.p_N(n) \\
&= 1.\frac{1}{10} + 2.\frac{2}{10} + 3.\frac{3}{10} + 4.\frac{4}{10} \\
&= 3 \\
V[N] &= E[N^2] - E[N]^2 \\
E[N^2] &= \sum_{n \in N} n^2 p_N(n) \\
&= 1.\frac{1}{10} + 2.\frac{2^3}{10} + 3.\frac{3^3}{10} + 4.\frac{4^3}{10} \\
&= 10 \\
\text{Therefore, } V[N] &= 10 - 9 \\
&= 1
\end{aligned}
$$

□

b).

$$
\begin{aligned}
p_{N/X}(n/x) &= \frac{p_{X/N}(x/n).p_N(n)}{p_X(x)} \\
p(x=0) &= \sum_{n\in N} p(x=0,n) \\
&= \sum_{n\in N} p(x=0/n).p(n) \\
&= p(x=0/n=1).p(n=1) + p(x=0/n=2).p(n=2) \\
&+ p(x=0/n=3).p(n=3) + p(x=0/n=4).p(n=4) \\
&= 0.\frac{1}{10} + \frac{1}{4}.\frac{2}{10} + \frac{1}{3}.\frac{3}{10} + \frac{3}{8}.\frac{4}{10} \\
&= \frac{3}{10}
\end{aligned}
$$

Since $X \in \{0,1\}$

$$
\begin{aligned}
p(x=1) &= 1 - p(x=0) \\
&= \frac{7}{10}
\end{aligned}
$$

For $X = 0$,

$$
\begin{aligned}
p(n=1/x=0) &= \frac{p(x=0/n=1).p(n=1)}{p(x=0)} \\
&= \frac{\frac{0x1}{10}}{\frac{3}{10}} \\
&= 0 \\
p(n=2/x=0) &= \frac{p(x=0/n=2).p(n=2)}{p(x=0)} \\
&= \frac{\frac{1x2}{4x10}}{\frac{3}{10}} \\
&= \frac{1}{6} \\
p(n=3/x=0) &= \frac{p(x=0/n=3).p(n=3)}{p(x=0)} \\
&= \frac{\frac{1x3}{3x10}}{\frac{3}{10}} \\
&= \frac{1}{3} \\
p(n=4/x=0) &= \frac{p(x=0/n=4).p(n=4)}{p(x=0)} \\
&= \frac{\frac{3x4}{8x10}}{\frac{3}{10}} \\
&= \frac{1}{2}
\end{aligned}
$$

For $X = 1$,

$$\begin{aligned}
p(n=1/x=1) &= \frac{p(x=1/n=1).p(n=1)}{p(x=1)} \\
&= \frac{1x\frac{1}{10}}{\frac{7}{10}} \\
&= \frac{1}{7} \\
p(n=2/x=1) &= \frac{p(x=1/n=2).p(n=2)}{p(x=1)} \\
&= \frac{\frac{3x2}{4x10}}{\frac{7}{10}} \\
&= \frac{3}{14} \\
p(n=3/x=1) &= \frac{p(x=1/n=3).p(n=3)}{p(x=1)} \\
&= \frac{\frac{2x3}{3x10}}{\frac{7}{10}} \\
&= \frac{2}{7} \\
p(n=4/x=1) &= \frac{p(x=1/n=4).p(n=4)}{p(x=1)} \\
&= \frac{\frac{5x4}{8x10}}{\frac{7}{10}} \\
&= \frac{5}{14}
\end{aligned}$$

$\square$

n/x	n=1	n=2	n=3	n=4
x=0	0	1/6	1/3	1/2
x=1	1/7	3/14	2/7	5/14

c).

$$\begin{aligned}
E[N|X=1] &= \sum_{n\in N} n.p(N=n|X=1) \\
&= 1.p(N=1|X=1) + 2.p(N=2|X=1) + 3.p(N=3|X=1) + 4.p(N=4|X=1) \\
&= 1.\frac{1}{7} + 2.\frac{3}{14} + 3.\frac{2}{7} + 4..\frac{5}{14} \\
&= \frac{20}{7}
\end{aligned}$$

$\square$

Ans 6. *a).*

$$
\begin{aligned}
f_{Y|X}(Y|X) &= \frac{f_{XY}(x,y)}{f_X(x)} \\
\text{Now, } f_X(x) &= \int_0^{4-2x} \frac{3}{16}(4-2x-y)dy \\
&= \frac{3}{16}[(4-2x)\int_0^{4-2x} 1dy + \int_0^{4-2x} ydy] \\
&= \frac{3}{16}[(4-2x)^2 - \frac{(4-2x)^2}{2}] \\
&= \frac{3*(4-2x)^2}{16*2} \\
&= \frac{3*4*(2-x)^2}{32} \\
&= \frac{3*(2-x)^2}{8} \\
\text{Therefore, } f_{Y|X}(Y|X) &= \frac{\frac{3}{16}(4-2x-y)}{\frac{3}{8}(2-x)^2} \\
&= \frac{1}{2}.\frac{4-2x-y}{(2-x)^2}
\end{aligned}
$$

□

b).

$$
\begin{aligned}
P(Y \geq 2|X=1/2) &= \int_2^3 \frac{3-y}{2(2-\frac{1}{2})^2}dy \\
&= \int_2^3 \frac{3-y}{\frac{2*9}{4}} \\
&= \frac{2}{9}\int_2^3 (3-y)dy \\
&= \frac{2}{9}(3\int_2^3 1dy - \int_2^3 ydy) \\
&= \frac{2}{9}(3(1) - \frac{9-4}{2}) \\
&= \frac{1}{9}
\end{aligned}
$$

□

c).

$$
\begin{aligned}
E(Y|X) &= \int_0^{4-2x} Y.F(Y|X)dy \\
&= \int_0^{4-2x} y.\frac{1}{2}.\frac{4-2x-y}{(2-x)^2}dy \\
&= \frac{1}{2.(2-x)^2}\int_0^{4-2x}(4-2x).y - y^2dy \\
&= \frac{1}{2.(2-x)^2}((4-2x)\int_0^{4-2x} ydy - \int_0^{4-2x} y^2dy) \\
&= \frac{1}{2.(2-x)^2}(\frac{(4-2x).(4-2x)^2}{2} - \frac{(4-2x)^3}{3}) \\
&= \frac{1}{2.(2-x)^2}\frac{(4-2x)^3}{12(2-x)^2} \\
&= \frac{8*(2-x)^3}{12*(2-x)^2} \\
&= \frac{2}{3}(2-x)
\end{aligned}
$$

□

Ans 7.

a) 68% values lie within 1 standard deviation. According to the Central limit theorem, when we increase number of samples we get answer nearer to the actual mean. Therefore, as we increase the number of samples in simulate.py we get the answer nearer to 0.
Now, 99.7 values lie within 3 standard deviation. So when we set the standard deviation to 10 we get a lot more varied samples in comparison to standard deviation = 1. Therefore, we get the answer which is not as close to the actual mean in comparison to s.d=1
b)For the co-variance $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ the samples of the normal distribution have freedom in all 3 Dimensions so they scatter around the 3D space
When we change the co-variance to $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, all points lie in the x-z=0 plane since x = 1 and z =1 and y =0